⑲日本国特許庁（JP）　⑪特許出願公開

⑫公開特許公報（A）　平2－28879

⑤Int. Cl.⁵	識別記号	庁内整理番号
G 06 F 15/40	530 S	7313－5B
C 07 B 61/00	Z	7457－4H
G 06 F 15/42	M	7313－5B

⑬公開　平成2年(1990)1月30日

審査請求　未請求　請求項の数 1　(全6頁)

⑤発明の名称　化学構造式の完全一致検索方式

⑳特　願　昭63－179698
㉒出　願　昭63(1988)7月19日

⑫発明者　溝部　吉巳　東京都港区芝5丁目33番1号　日本電気株式会社内
⑫発明者　吉田　元二　大阪府大阪市此花区春日出中3丁目1番98号　住友化学工業株式会社大阪研究所内
⑪出願人　日本電気株式会社　東京都港区芝5丁目33番1号
⑪出願人　住友化学工業株式会社　大阪府大阪市東区北浜5丁目15番地
⑭代理人　弁理士　境　廣巳

明　細　書

1.発明の名称

化学構造式の完全一致検索方式

2.特許請求の範囲

検索する化学構造式から分子式等を生成する分子式等生成部と、

化学構造式の中に特定の部分構造が存在するかどうかを示すフラグメントコード列を抽出する部分構造抽出部と、

生成，抽出した分子式，フラグメントコード列等を用いてデータの索引を探す検索部と、

検索されたデータを1件単位で読み取って原子を1対1で対応させて同じものかどうかを判定する構造判定部とを有してなることを特徴とした化学構造式の完全一致検索方式。

3.発明の詳細な説明

〔産業上の利用分野〕

本発明は多量の化合物データの中から特定の化合物を検索する方式に関するものであり、特に化学構造式が完全に一致するものを高速に検索する方式に関する。

〔従来の技術〕

従来、この種の化合物検索は化合物の構造をマトリックスで表現したコネクションテーブル（CT）と呼ばれるデータを作成し、このコネクションテーブルに対して配列演算を繰り返し、一意な配列に変換してそれを記号列で表し、蓄積された化合物データの記号列と同じかどうかを比較・判別することにより化学構造式を特定するようにしていた。

すなわち、第5図(a)の如き化学構造式のコネクションテーブルを作成する場合、第5図(b)のように各原子に①，②，……というように順に番号を付け、第5図(c)のように各原子間の結合状態を示せばよい。なお、＊は1重結合、＊＊は2重結合を示している。そして、第5図(c)のコネクションテーブルに配列演算を繰り返し、第5図(e)に示す一意なコネクションテーブルを見つけ出す。これは、第5図(b)の各原子の番号を第5図(d)のように付け変えることと同じである。そして、第5図(e)

のコネクションテーブルを所定の規則に従って記号列にし、比較用のデータを得ていた。

〔発明が解決しようとする課題〕

ところで、上述した従来の化合物検索にあっては、化学構造式に対応させて自由な原子の番号付けに基づいて作成したコネクションテーブルから一意な番号付けのコネクションテーブルを見つけ出す作業が必要であり、その組み合わせは原子の数をnとすると最大でn！組だけあることから、原子の数が多くなると長時間の演算が必要になり、検索全体に要する時間が長くなるという欠点があった。

本発明は上記の点に鑑み提案されたものであり、その目的とするところは、高速に検索が行える化学構造式の完全一致検索方式を提供することにある。

〔課題を解決するための手段〕

本発明は上記の目的を達成するため、検索する化学構造式から分子式等を生成する分子式等生成部と、化学構造式の中に特定の部分構造が存在するかどうかを示すフラグメントコード列を抽出する部分構造抽出部と、生成，抽出した分子式，フラグメントコード列等を用いてデータの索引を探す検索部と、検索されたデータを1件単位で読み取って原子を1対1で対応させて同じものかどうかを判定する構造判定部とを有している。

〔作用〕

本発明の化学構造式の完全一致検索方式にあっては、検索する化学構造式から分子式等生成部が分子式等を生成し、部分構造抽出部が化学構造式の中に特定の部分構造が存在するかどうかを示すフラグメントコード列を抽出し、これらの生成，抽出された分子式，フラグメントコード列等を用いて検索部がデータの索引を探し、候補が絞られた状態で構造判定部が検索されたデータを1件単位で読み取り、原子を1対1で対応させて同じものかどうかを判定し、化合物を特定する。

〔実施例〕

次に、本発明の実施例について図面を参照して説明する。

3

第1図は本発明の化学構造式の完全一致検索方式の一実施例を示す構成図であり、機能部（太枠で示すブロック）とデータおよび処理の流れをいっしょに示してある。第1図において、分子式等抽出部1は検索する化学構造式から原子，ボンド（結合）の種別，数等を抽出して分子式や識別コードを生成する機能を有している。また、部分構造抽出部2は色々な部分構造を示すフラグメントコード列を化学構造式から抽出する機能を有している。そして、検索部3は分子式等抽出部1および部分構造抽出部2により生成，抽出された分子式，識別コード，フラグメントコード列等を索引として索引データの中から該当するものを探す機能を有し、構造判定部4は検索部3で見つけた類似化学構造式にかかるデータを1件単位で読み込み、原子を1つずつ対応させて同じものかどうかを詳細に判定して最終的に化合物を特定する機能を有している。

以下、具体例を交えて動作を説明する。

第1図において、検索する化学構造式が例えば

4

コネクションテーブルの形で与えられると、分子式等抽出部1は化学構造式から原子種別毎の数を求め、原子名とその数を所定の規則で並べて分子式（例えばC4H6）を生成する。なお、同じ分子式であっても異なる化学構造式がいくつか存在することもあるので、ボンド種別毎の数等を並べた識別コードを生成し、更に細かく識別できるようにする。第2図は分子式がC4H6である化学構造式の例を(a)～(g)の如く7つ示し、各々の化学構造式に対応して識別コードとしての結合数列を示してある。なお、結合数列は以下の数をスラッシュ「／」で区切って順番に並べたものである。

- 2重ボンドの数
- 3重ボンドの数
- 環内の1重ボンドの数
- 環内の2重ボンドの数
- 結合相手が3つの原子の数
- 結合相手が4つの原子の数

第2図の例で示す識別コードでは、(c)と(d)とは同じ識別コード(0/1/0/0/0/2)になり識別できない

5

6

が、他の5つは完全に1つを識別することができる。なお、第2図の識別コードに＋イオンや－イオンの数等を付加すると、更に細かく識別することができる。

次いで、第1図において、検索部3は分子式等抽出部1で作られた分子式，識別コードを使用して索引データから検索を行い、通常はこの段階で該当する化合物が複数検索される。なお、この検索段階で該当する化合物が0件であれば、同一化学構造式は蓄積されたデータ中にはないと判定でき、検索は終了する。また、該当する化合物が1件の場合は、次の部分構造抽出部2および検索部3による処理を経由せずに、直接に構造判定部4の処理に移行させるようにしてもよい。

さて、分子式等抽出部1の分子式，識別コードを用いた検索で該当する化合物が検索された場合、処理は部分構造抽出部2に進み、コード中の各文字が部分構造の有無または数を示すフラグメントコード列が抽出される。すなわち、分子式やボンドの数等の情報からなる識別コードだけでは第2

図の(c)と同のように識別できない場合があるので、より細かく識別するためにフラグメントコード列が作成される。第3図はフラグメントコード列の例を示しており、各文字「i」は対応する部分構造の数を示し、例えばi＝0の場合は該当する部分　造がないことを示し、i＝1の場合は該当する部分構造が1個存在することを示し、i＝nの場合は該当する部分構造がn個以上存在することを示す。なお、通常はnを1として簡略化する。そして、化学構造式の識別のためにフラグメントコード列全体を1つの識別コードとして扱う。

次いで、第1図において、検索部3は部分構造抽出部2で作られたフラグメントコード列を使用して、前回の分子式等抽出部1の分子式，識別コードにより検索された複数の索引データから再度検索を行い、該当する化合物を検索する。なお、この検索段階で該当する化合物が0件であれば、同一化学構造式は蓄積されたデータ中にはないと判定でき、検索は終了する。なお、分子式等抽出部1の分子式，識別コードによる検索と部分構造

抽出部2のフラグメントコード列による検索とは図示の例に限られず、2～3段階に分けて行うことも可能である。

しかして、検索部3で最後まで類似化学構造式であると判定されたものは通常数件以下になり、構造判定部4はそれらの化学構造式のデータを1件単位で読み取り、1原子単位で対応を取って同一化合物かどうかを最終的に判定し、化合物を特定する。

このように、予め分子式，識別コード，フラグメントコード列等を用いて蓄積された化合物データの中から候補を絞り、それらに対して1原子単位で比較を行うため、高速に正確な検索が行えるものである。また、第6図(a)，(b)に示すように化学構造式は異なるが同じ化合物（この例は芳香環）を表すものや、異性体，同位体原子，配位結合等を含む化合物のように1つの化合物で2通りの書き方があるもの、および、第6図(c)のように1つの書き方で第6図(d)，(e)のような2つの化合物の混合物を表すもの等は、1つのものを必ず一意な

記号列にしなければならない従来の化合物検索では取り扱えないという欠点があったが、本発明では最終的に1原子単位で比較を行うため、そのような不都合はない。

次に、第4図は本発明の応用例を示したものあり、新しい化学構造式をデータとして蓄積する時に、同一化合物が蓄積済みであることによる2重蓄積を防ぐようにしたものである。なお、第4図における分子式等抽出部1，部分構造抽出部2，検索部3，構造判定部4は前述の実施例と同じものであり、検索部3は分子式等抽出部1と部分構造抽出部2に対して共通して示してあるが動作は同様である。新たな構成としては、構造式蓄積部5と索引蓄積部6とが加わっている。

動作にあっては、分子式等抽出部1と部分構造抽出部2は、蓄積しようとする化合物から索引データを作成するため、および既に蓄積済みの同一化合物があるかどうかを調べるために、分子式，フラグメントコード列等を抽出する。検索部3は分子式やフラグメントコード列等を用いて索引デ

ータを探し、該当する化合物が見つからなければよいが、見つかった場合は完全に同じものかどうかを構造判定部4で判定する。そして、同じものがなければ　造式蓄積部5は新規の化合物として構造判定部4で使用したその化合物の構造式データを蓄積する。次いで、索引蓄積部6はこの化合物の索引データとして、分子式等抽出部1と部分構造抽出部2とで抽出した分子式，フラグメントコード列等を蓄積する。なお、これらの蓄積された構造式データと索引データは、その後の検索や蓄積時の2重登録チェックで使用される。

〔発明の効果〕

以上説明したように、本発明の化学構造式の完全一致検索方式にあっては、分子式等抽出部と部分構造抽出部とで生成，抽出される分子式，識別コード，フラグメントコード列等を用いて大まかな検索を行い、次いで該当するデータから1原子単位で判定を行うため、正確な検索が高速で行える効果がある。

4.図面の簡単な説明

第1図は本発明の化学構造式の完全一致検索方式の一実施例を示す構成図、

第2図は分子式等抽出部で生成される分子式，識別コードの例を示す図、

第3図は部分構造抽出部で抽出されるフラグメントコード列の例を示す図、

第4図は本発明の応用例を示す図、

第5図はコネクションテーブルの例を示す図および、

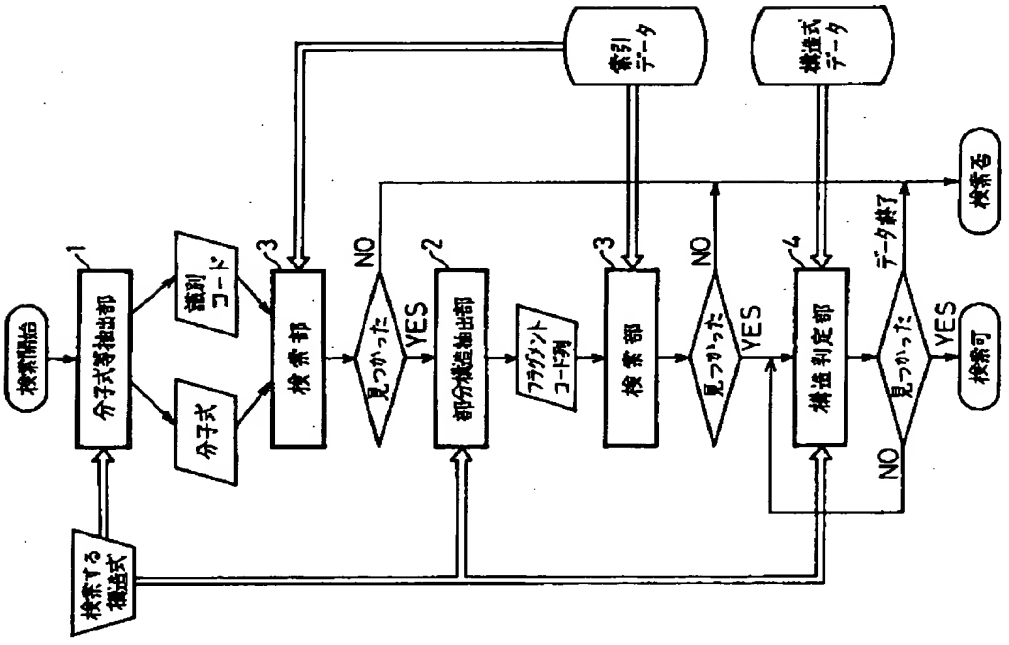
第6図は従来では識別できなかった化学構造式の例を示す図である。

図において、1…分子式等抽出部、2…部分構造抽出部、3…検索部、4…構造判定部、5…構造式蓄積部、6…索引蓄積部

特許出願人　日本電気株式会社　外1名

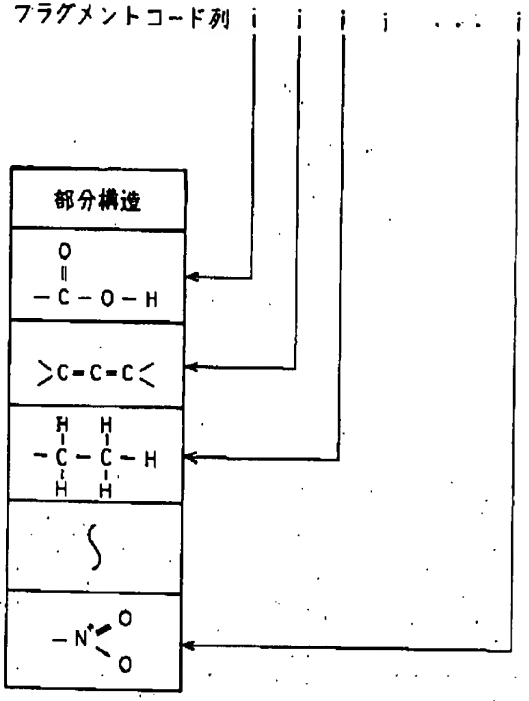
代理人　弁理士　境　　廣　巳

11　　12

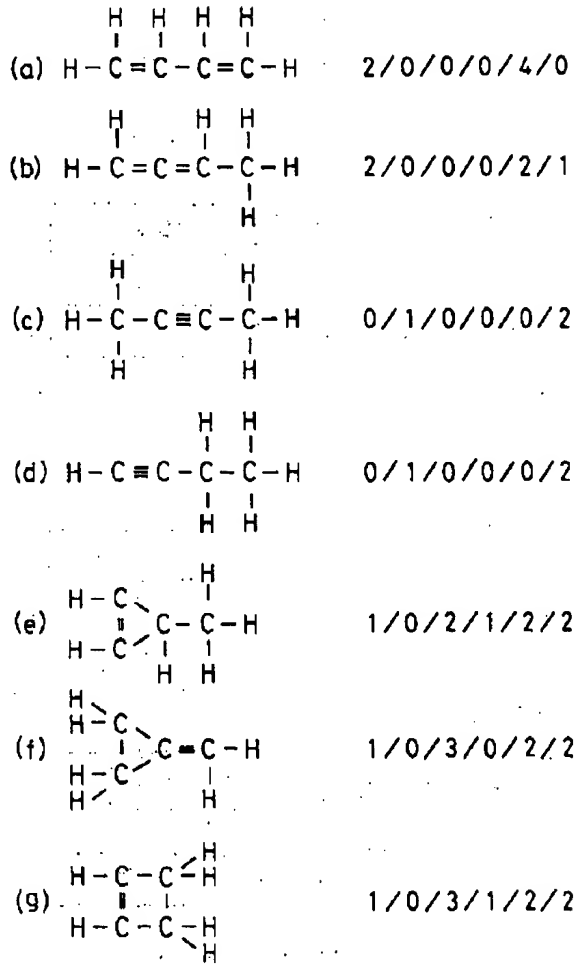


本発明の一実施例の構成図

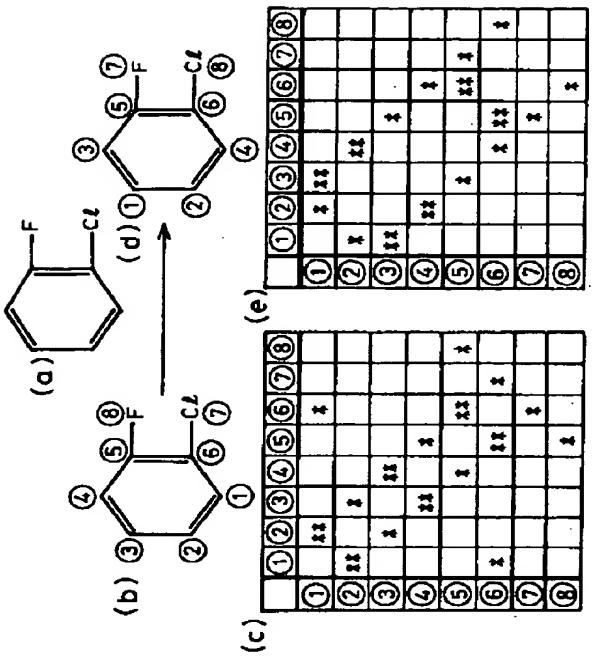
第1図



フラグメントコード列の説明図
第 3 図

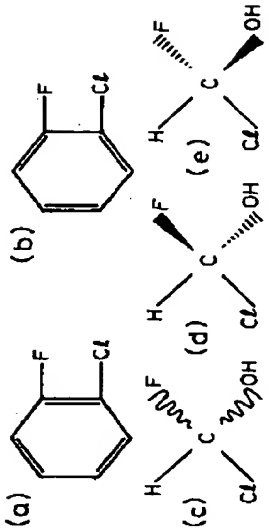


分子式，識別コードの説明図
第 2 図



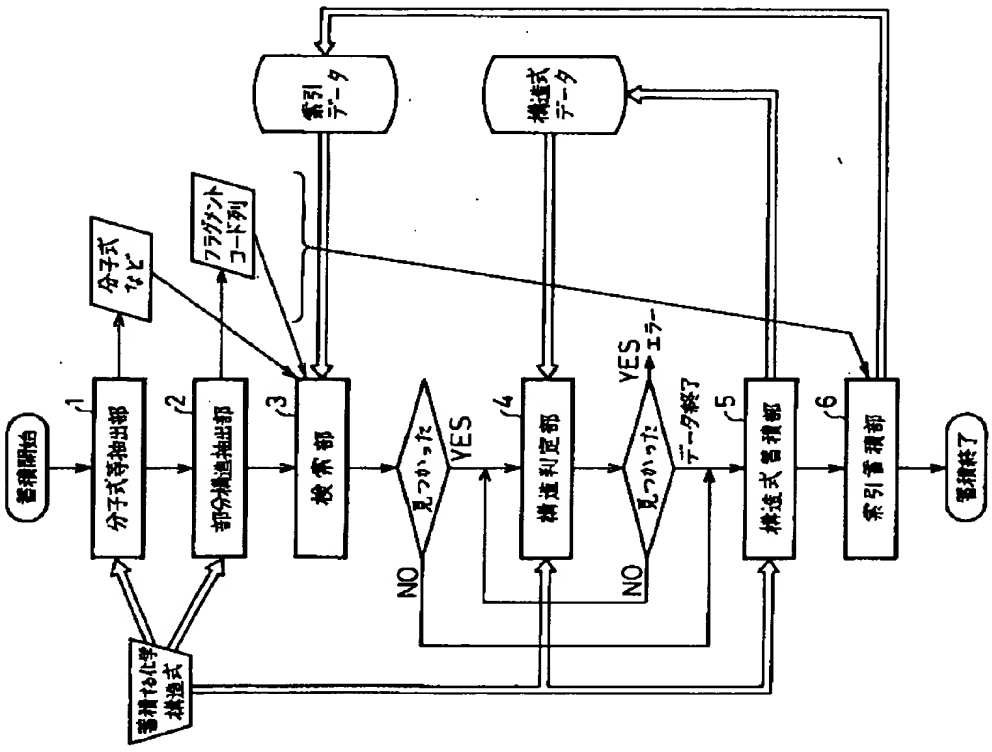
コネクションテーブルの説明図

第 5 図



従来において識別できなかった化学構造式の例

第 6 図



本発明の応用例の構成例

第 4 図